Identification and characterization of a novel chitinase-like gene cluster (AgCht5) possibly derived from tandem duplications in the African malaria mosquito, Anopheles gambiae

Jianzhen Zhang, Xin Zhang, Yasuyuki Arakane, Subbaratnam Muthukrishnan, Karl J. Kramer, Enbo Ma, Kun Yan Zhu

Published Version Information

Citation: Zhang, J., Zhang, X., Arakane, Y., Muthukrishnan, S., Ma, E., Kramer, K.J., Zhu, K.Y. (2011). Identification and characterization of a novel chitinase-like gene cluster (AgCht5) possibly derived from tandem duplications in the African malaria mosquito, Anopheles gambiae. Insect Biochemistry and Molecular Biology, 41(8), 521-528.

Digital Object Identifier (DOI): doi:10.1016/j.ibmb.2011.03.001

Publisher's Link: www.elsevier.com/locate/ibmb

Submitted to: **Insect Biochemistry and Molecular Biology** **MS# IB-D-11-00002**

Identification and characterization of a novel chitinase-like gene cluster (*AgCht5*) possibly derived from tandem duplications in the African malaria mosquito, *Anopheles gambiae*

Jianzhen Zhang [a,b], Xin Zhang [b], Yasuyuki Arakane [c, d], Subbaratnam Muthukrishnan [c], Karl J. Kramer [c], Enbo Ma [a] and Kun Yan Zhu [b,*]

[a] Research Institute of Applied Biology, Shanxi University, Taiyuan, Shanxi 030006, China

[b] Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS 66506, USA

[c] Department of Biochemistry, 141 Chalmers Hall, Kansas State University, Manhattan KS 66506, USA

[d] Division of Plant Biotechnology, College of Agriculture and Life Science, Chonnam National University, Gwangju 500-757, Korea

* Corresponding author: Tel.: +1 785 532 4721; fax: +1 785 532 6232.

E-mail: kzhu@ksu.edu (K.Y. Zhu)

ABSTRACT

Insect chitinase 5 (Cht5), a well-characterized enzyme found in the molting fluid and/or integument, is classified as a group I chitinase and is usually encoded by a single gene. In this study, a *Cht5* gene cluster consisting of five different chitinase-like genes (*AgCht5-1*, *AgCht5-2*, *AgCht5-3*, *AgCht5-4* and *AgCht5-5*) was identified by a bioinformatics search of the genome of *Anopheles gambiae.* The gene models were confirmed by cloning and sequencing of the corresponding cDNAs and gene expression profiles during insect development were determined. All of these genes are found in a single cluster on chromosome 2R. Their open reading frames (ORF) range from 1227 to 1713 bp capable of encoding putative proteins ranging in size from 409 to 571 amino acids. The identities of their cDNA sequences range from 52 to 66%, and the identities of their deduced amino acid sequences range from 38 to 53%. There are four introns for *AgCht5-1*, two for *AgCht5-2* and *AgCht5-3*, only one for *AgCht5-4*, but none for *AgCht5-5* in the genome. All five chitinase-like proteins possess a catalytic domain with all of the conserved sequence motifs, but only AgCht5-1 has a chitin-binding domain. Phylogenetic analysis of these deduced proteins along with those from other insect species suggests that AgCht5-1 is orthologous to the Cht5 proteins identified in other insect species. The differences in expression patterns of these genes at different developmental stages further support that these genes may have distinct functions. Additional searching of the genomes of two other mosquito species led to the discovery of four *Cht5*-like genes in *Aedes aegypti* and three in *Culex quinquefasciatus*. Thus, the presence of a *Cht5* gene cluster appears to be unique to mosquito species and these genes may have resulted from gene tandem duplications.

Keywords: African malaria mosquito, *Anopheles gambiae*, Chitinase, Gene cluster, Gene duplication, Glycoside hydrolase

1. Introduction

Chitinases (EC.3.2.1.14) are enzymes responsible for hydrolyzing glycosidic bonds in chitin and widely distributed in nature, including vertebrates, microorganism and even plants. Mammals are not known to synthesize chitin or metabolize chitin as a nutrient, yet the human genome encodes eight different chitinases of the glycoside hydrolase 18 (GH18) family, which play an important role in T-cell mediated inflammation and asthma (Funkhouser et al., 2007; Reese et al., 2007; Shuhui et al., 2009).

In insects, chitin associates with proteins to form the cuticular exoskeleton and peritrophic matrix (PM) in the midgut lumen. During a molting cycle, a part of the old cuticle is digested while new chitin is synthesized and deposited (Reynolds et al., 1996). It has been suggested that insect chitinases may have multiple functions including defense, digestion and molting (Shen and Jacob-Lorena, 1997; Filho et al., 2002; Zheng et al., 2002; Genta et al., 2006; Zhu et al., 2008b). Indeed, a chitinase expressed in the gut of European corn borer (*Ostrinia nubilalis*) has been identified and it has been proposed that this enzyme is responsible for regulation of chitin content of PM and growth of *O. nubilalis* larvae (Khajuria et al., 2010). A chitinase was also purified from the venom gland of an endoparasitic wasp *Chelonus* sp. near *curuimaculatus* (Krishnan et al., 1994). Because of the crucial roles of chitinases in insect growth and development, these enzymes have been widely recognized as potential targets for developing chemical pesticides for insect control (Royer et al., 2002; Hirose et al., 2010).

Insect chitinases belong to the GH18 multi-gene family with a rapid increase in the number of genes identified as the annotation of completed genome sequences of several

insect species has occurred (Zhu et al., 2004, 2008a). At present, insect chitinases and chitinase-like proteins are classified into eight groups based on a phylogenetic analysis of their catalytic domains (Arakane and Muthukrishnan, 2010). Among these chitinases and chitinase-like proteins, chitinase 5 (Cht5) is classified into group I.

To date, only a single *Cht5* gene has been identified from each of several insect genomes and the representatives of this gene have been well characterized in several lepidopteran and coleopteran species (Kim et al.,1998; Shinoda et al., 2001; Zheng et al., 2002; Ahmad et al., 2003; Fitches et al., 2004; Bolognesi et al., 2005; Zhu et al., 2008a). All insect Cht5s have a typical multidomain structural organization that includes a signal peptide, a catalytic domain, a PEST-like linker region enriched in proline (P), glutamic acid (E), serine (S) and threonine (T) that is heavily glycosylated, and a cysteine-rich chitin-binding domain. The sizes of Cht5 enzymes range from 552 to 586 amino acid residues. The transcripts of *Cht5* are mainly detected in the epidermis and the gut, and its expression increases during the molting process (Royer et al., 2002). In *Helicoverpa armigera*, *Cht5* is also expressed in the fatbodies (Ahmad et al., 2003). However, the expression of *Cht5* has not been reported in haemocytes from any insect species. In *Manduca sexta*, the transcript level of *Cht5* can be induced by 20-hydroxyecdysone, but it is suppressed by the juvenile hormone analog, fenoxycarb (Kramer et al., 1993). The recombinant protein expressed in an insect cell line showed high levels of chitinolytic activity (Gopalakrishnan et al., 1995; Zheng et al., 2003; Ahmad et al., 2003).

It is likely that Cht5 may be involved in chitin turnover associated with molting. In *Tribolium castaneum*, RNA interference was performed to silence *TcCht5*. The insects that

were injected with double-stranded RNA (dsRNA) for *TcCht5* exhibited a lethal phenotype only at the pharate adult stage. At the time of death, some of the adult cuticle was visible under the old pupal cuticle which was not shed, suggesting that *TcCht5* is required for pupal–adult molting (Zhu et al., 2008b). All these results indicate that Cht5 is an essential enzyme for insect growth and development.

Although insects have been known to have only a single Cht5 gene, our recent studies have revealed a novel Cht5 gene cluster consisting of multiple chitinase-like genes in three mosquito species. In this paper, we report: 1) identification of a cluster of five *An. gambiae Cht5*-like genes (*AgCht5-1*, *AgCht5-2*, *AgCht5-3*, *AgCht5-4* and *AgCht5-5*) and their chromosomal localization, 2) characterizations of their gene models and developmental expression patterns, and 3) the results of a comparative investigation on *Cht5* gene clusters in two other mosquito species including *Aedes aegypti* and *Culex quinquefasciatus*. This is the first demonstration of gene duplication of this group of chitinase genes, which may be unique to the mosquito lineage.

2. Materials and Methods

2.1. Insect culture

An. gambiae was obtained from the Malaria Research and Reference Reagent Resource Center (MR4, Manassas, VA) and has been cultured in the Department of Entomology at Kansas State University, Manhattan, KS since 2005. The colony was maintained based on the methods previously described (Zhang and Zhu, 2006)

2.2. Sequencing of cDNAs of chitinase 5 gene cluster

A bioinformatics search was conducted to identify different chitinase and chitinase-like genes in the genome of *An. gambiae*. Based on the bioinformatics analysis, we identified a chitinase 5-like gene (*AgCht5*) (accession no: XP_001237469.2) that has been annotated as one encoding a large protein with five chitinase catalytic domains. Each chitinase-coding domain of this gene was then searched against the *An. gambiae* EST database. Individual EST clones obtained from the MR4 were sequenced. For the EST clones missing the 3'-end sequences of *AgCht5-1* and *AgCht5-4*, 3'-RACE PCR was performed to obtain their full-length cDNA by using the SMART™ RACE cDNA Amplification Kit (Clontech, Mountain View, CA). The primer sequences for 3'-RACE PCR are shown in Table 1. The PCR products were subcloned into the TA cloning vector (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol and sequenced by DNA Sequencing Facility at Kansas State University (Manhattan, KS).

Multiple sequence alignment was performed using NPS@ (Network Protein Sequence Analysis). Identity comparisons among the five AgCht5 sequences were performed by using DNAstar (Madison, WI). SMART domain analysis (http://smart.embl-heidelberg.de/; Schultz et al., 1998) and UCSC genome bioinformatics programs (http://genome.ucsc.edu) were used to predict the domain architecture and gene structure of each identified chitinase and chitinase-like gene, respectively. The phylogenetic tree was constructed based on the amino acid sequences of their catalytic domains by the neighbor-joining algorithm using Mega 4.0 software (Tamura et al., 2007).

2.3. Analysis of developmental stage-dependent gene expression patterns

The expression patterns of the five *AgCht5* genes at different developmental stages including eggs; first-, second-, third- and fourth-instar larvae; and adults of *An. gambiae* were evaluated in the study. For more detailed developmental expression patterns, mosquito eggs and pupae were collected at several time points for each developmental stage. Total RNA was isolated using the Trizol reagent (Invitrogen) and treated with DNase Ⅰ (Fermentas, Glen Burnie, MD). The first-strand cDNA was synthesized using a First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions. Beacon 7.0 software was used for primer design and ribosomal protein S3 (*Rps3*) was used as an internal reference gene. The primers used for expression analysis are shown in Table 1. RT-PCR was carried out in a 25-μl reaction mixture containing 1 μl template cDNA, 12.5 μl Taq Master Mix (Fermentas), 0.2 μM of each primer and sterilized water. The thermal cycle program for RT-PCR consisted of an initial denaturation at 94°C for 1 min followed by 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 45s, and a final extension at 72°C for 5 min. PCR products were analyzed on a 2% agarose gel. Three biological replications (i.e., 3 independent preparations of total RNA), each with three repeated PCR runs, were performed in this analysis.

3. Results

3.1. Full-length cDNAs and the deduced amino acid sequences of five AgCht5 *genes*

The conceptual translation of the *An. gambiae* gene model XP_001237469.2 predicts a 2095 amino acid-long protein with five catalytic domains. However, analysis of the sequences of five cDNA clones (EST clone numbers: 19600449629438 for *AgCht5-1*, 19600449653107 for *AgCht5-2*, 19600449656904 for *AgCht5-3*, NAP1-P158-B-06-5 for *AgCht5-4*, and 19600449684410 for *AgCht5-5*) that were obtained from the MR4 failed to

provide evidence for a long transcript that bridges genomic sequences encoding adjacent chitinase catalytic domains predicted by this gene model. Instead, we could detect sequences that were presumed to be introns in this gene model at the 5'ends or 3'ends of the five full-length chitinase cDNA sequences that we have characterized.

The additional sequences at the 3'-ends of these clones (which were not included in the gene model XP_001237469.2) had stop codons and polyadenylation signal sequences and short poly A tails. The additional sequences at the 5'-ends of the full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones are consistent with a gene model that replaces the current XP_001237469.2 with five separate genes encoding five chitinases that differ in their leader peptide as well as catalytic domain sequences. Each of the five cDNA sequences contains a start codon (ATG) and a stop codon (TAA, TAG, or TGA) as well as a poly (A) tail. Except for *AgCht5-4* that lacks a typical polyadenylation signal sequence (AATAAA), all the remaining four cDNAs contain such a signal sequence. These five chitinase-like genes are denoted as *AgCht5-1*, *AgCht5-2*, *AgCht5-3*, *AgCht5-4* and *AgCht5-5*. Their cDNA and deduced amino acid sequences have been deposited in GenBank with the following accession numbers: HQ456129 for *AgCht5-1*, HQ456130 for *AgCht5-2*, HQ456131 for *AgCht5-3*, HQ456132 for *AgCht5-4* and HQ456133 for *AgCht5-5*.

Analysis of the genomic organization of the five *AgCht5* genes showed that they form a contiguous cluster of genes in chromosome 2R (Fig. 1A, Table 2). The shortest distance between two of these genes is only 340 bp, whereas the longest distance is 2045 bp. The percent nucleotide sequence identities among the cDNAs of the five genes range from 52 to

66% (Table 3). *AgCht5-1* has four introns, *AgCht5-2* and *AgCht5-3* have two introns, *AgCht5-4* has only one intron, whereas *AgCht5-5* has no introns (Fig. 1B).

The five *AgCht5* genes were predicted to encode five chitinase-like proteins with sizes ranging from 409 to 571 amino acid residues (Fig. 1C). The identities of the amino acid sequences among the five full-length deduced proteins range from 38 to 53% (Table 3). All of the five putative chitinase proteins possess a catalytic domain, but only AgCht5-1 exhibits a chitin-binding domain (Fig. 2). The catalytic domain of each deduced protein is composed of four motif sequences that are conserved among family 18 chitinases. All the five deduced proteins were predicted to possess a signal peptide (Fig. 1C, Fig. 2).

To examine whether similar *Cht5* gene clusters exist in other mosquito species, we searched for *Cht5*-related genes in the genome databases of *Ae. aegypti* and *C. quinquefasciatus*, and identified four *AaCht5* and three *CqCht5* genes, respectively. By aligning the deduced amino acid sequences of all of the Cht5s from the three mosquito species and other insect species, we assigned the names *AaCht5-1*, *AaCht5-2*, *AaCht5-3* and *AaCht5-4* for those identified in *Ae. aegypti*, and *CqCht5-1*, *CqCht5-2* and *CqCht3* for those identified in *C. quinquefasciatus*. Analysis of their domain architectures indicated that all the deduced Cht5 proteins from the three mosquito species have a catalytic domain, but only the first Cht5 protein in each species (i.e., *AgCht5-1*, *AaCht5-1*and *CqCht5-1*) contains a chitin-binding domain (Fig. 3).

3.2. Phylogenetic analysis of five deduced AgCht5 protein sequences

To explore the relationship among the insect Cht5s, a phylogenetic tree was constructed based on the sequences of their catalytic domains. Results showed that all of the insect Cht5s

fall into two branches supported by a bootstrap value of 100 after 5000 replications (Fig. 4). The first group represents the mosquito Cht5-1 and all other well characterized insect Cht5s with chitinase activities, whereas the second group represents the remaining mosquito Cht5s. Apparently, AgCht5-1, AaCht5-1 and CqCht5-1 from the three mosquito species are more closely related and might represent mosquito orthologs of insect Cht5 enzymes. In contrast, the mosquito Cht5-2, Cht5-3, Cht5-4 and Cht5-5 are clustered in another branch that may be encoded by genes derived from an ancestral *Cht5-1* by gene duplications. Cht5-2s from three mosquito species close together with robust bootsrap value, suggest they are also orthologs.

3.3. Developmental stage-dependent expression patterns of five AgCht5 *genes*

The developmental stage-dependent expression patterns of different *AgCht5* transcripts were determined by RT-PCR. The levels of transcripts of the five *AgCht5* genes were apparently higher in third- and fourth-instars (Fig. 5A). Four genes including *AgCht-1*, *AgCht5-2*, *AgCht5-3* and *AgCht5-5* were expressed at all developmental stages, whereas *AgCht5-4* was expressed mainly in third- and fourth-instar larvae with trace amounts of transcripts detected in the eggs and first-instar larvae. Their detailed expression patterns were further examined in eggs collected at 12, 24, 36, 48 and 60 h after deposition by blood-fed females (Fig. 5B). High transcript levels were detected in 36-h eggs for all of the five *AgCht5s*. However, no detectable expression was found for *AgCht5-2* in 60-h eggs and for *AgCht5-5* in 12-h eggs. On the other hand, *AgCht5-4* was scarcely detected in mature eggs. Similarly, the expression patterns of the five different *AgCht5* genes were also examined in pupae collected at 0, 10, 20, 30 and 34 h after pupation (Fig. 5C). *AgCht5-1* and *AgCht5-3* were apparently expressed during all the pupal stage and exhibited similar expression patterns,

246 whereas *AgCht5-2* and *AgCht5-5* were mainly expressed in 0- and 10-h pupae. However, the
247 expression of *AgCht5-4* appeared to gradually increase with pupal development from 0 to 34
248 h.

249 **4. Discussion**

250 Currently, the deduced Cht5 proteins from different insect species are grouped into one
251 clade in the phylogenetic analysis of insect chitinases (Arakane and Muthukrishnan, 2010).
252 All chitnase5 proteins possess a typical multiple domain structural organization consisting of
253 a signal peptide, an N-terminal catalytic domain with four conserved motif sequences,
254 KXXXXXGGW, FDGXDLDWEYP, MXYDXXG and GXXXWXXDXD, a S/T-rich linker
255 region and a cysteine-rich chitin-binding domain that conforms to the consensus
256 C-(X_{11})-C-(X_5)- C-(X_9)- C-(X_{12})- C-(X_7)-C spacing of six cysteines that are predicted to
257 form three disulfide bonds. They are highly expressed in the epidermis, and the transcripts
258 appear just before ecdysis and disappear soon after ecdysis (Kramer, et al., 1993; Zheng et al.,
259 2002). Furthermore, results of RNA interference in *T. castaneum* suggested that Cht5 might
260 be involved in chitin turnover associated with molting (Zhu et al., 2008b). To date, however,
261 only single Cht5 gene has been reported in various insect species.

262 From our genome-wide searching of the *An. gambiae* genome, we putatively identified
263 a chitinase 5 gene (*AgCht5*) (accession no: XP_001237469.2) that has been annotated in
264 Vectorbase to encode a large protein with five different catalytic domains. However, our
265 careful studies unexpectedly revealed different gene expression patterns when we used
266 unique primer sets designed to amplify cDNAs from specific regions of *AgCht5*. This finding
267 prompted us to hypothesize that *AgCht5* actually is a gene cluster consisting of multiple

genes and the different catalytic domains might be encoded by different genes. To address this question, we utilized the cDNA sequence of each domain of AgCht5 to search the *An. gambiae* EST database to determine whether separate transcripts corresponding to the five catalytic domains existed in the EST database. As expected, we obtained five full-length cDNA sequences, each corresponding to only one of the five catalytic domains. No evidence for an EST with sequences from two or more adjoining Cht5 coding regions was obtained. Each of these five full-length cDNAs in the EST database apparently encode a chitinase containing only one catalytic domain with all four conserved motif sequences expected of chitinases (Fig. 2). As supported by the different expression patterns of these genes during different developmental stages of the mosquito, we conclude that the previously reported *An. gambiae Cht5* gene model actually represents a unique gene cluster consisting of five different chitinase or chitinase-like genes. To our knowledge, this is the first report on a multiple-member Cht5 gene cluster in insects.

Our further studies confirmed a similar clustering of *Cht5* genes in other mosquito species. Based on our genome search, we identified four *Cht5*-like genes in *Ae. aegypti* and three in *C. quinquefasciatus*. By aligning these mosquito Cht5 proteins with other known insect Cht5 proteins, we found that only one Cht5 catalytic domain (AgCht5-1, AaCht5-1 or CqCht5-1) from each mosquito species was clustered with other domains from known insect Cht5 proteins with a bootstrap value of 100 (Fig. 4), whereas the remaining Cht5s from these three mosquito species were grouped into a different cluster. These results suggest that *AgCht5-1*, *AaCht5-1*, *CqCht5-1* and all other known insect *Cht5s* are orthologous genes, whereas the remaining four *AgCht5*, three *AaCht5* and two *CqCht5* genes are paralogous to

290 *AgCht5-1*, *AaCht5-1* and *CqCht5-1*, respectively. AgCht5-2, AaCht5-2, CqCht5-2 were
291 clustered together with high bootstrap value, and represented similar domain architecture
292 (Fig.3), suggesting that *Cht5-2s* from three mosquito species may be also orthologous genes.
293 The absence of paralogs of Cht5 genes in *D. melanogaster* and *T. castaneum* suggests that
294 amplification of this subgroup of chitinase genes is of recent origin.
295 The catalytic and chitin-binding domains are two important structural components of
296 chitinases. Sequence motif analysis showed that all mosquito Cht5 proteins possess a
297 catalytic domain, but only three of the proteins, AgCht5-1, AaCht5-1 and CqCht5-1, contain
298 the signature sequence DWEYP within conserved region II, which is known to be located in
299 or near the catalytic site of the enzyme. The third residue E is crucial for a chitinase being
300 catalytically active because it probably serves as a proton donor in the catalytic mechanism
301 (Watanabe et al., 1994). Thus, AgCht5-1, AaCht5-1 and CqCht5-1 are predicted to be
302 catalytically active, which are similar with all the other known insect Cht5s. In contrast, all of
303 the other predicted Cht5-like proteins from the three mosquito species are likely to be
304 catalytically inactive because the E residue is replaced by L in these proteins. These proteins
305 may have carbohydrate-binding capability. Domain analysis revealed that only AgCht5-1,
306 AaCht5-1 and CqCht5-1 possess a chitin-binding domain. The function of chitin-binding
307 domain is presumably to anchor the enzyme tightly onto the large insoluble polymeric
308 substrate, thereby facilitating the hydrolytic process (Arakane et al., 2003, Boot et al., 2001).
309 Our analysis of the genomic organization of the *AgCht5* cluster suggests that this gene
310 cluster could be evolved from gene duplications. Gene duplication events are generally
311 considered to be essential in the evolution of gene families, which facilitate the generation of

312 new genes with new functions. Gene duplication can occur via three major mechanisms:
313 segmental duplication (of the whole genome, of one to a few chromosomes or of large parts
314 of a chromosome), tandem duplications (of one to a few adjacent genes), and retroposition or
315 other transposition events (Kong et al., 2007). Among these, tandem and segmental
316 duplication events contribute mostly to the generation of new members in nuclear gene
317 families. Tandem duplicates are copies of a nearby gene that are within short intron distances
318 of each other and may harbor some interesting biology. Gene expansion by tandem
319 duplication is common in cytochrome P450 gene evolution (Ai et al., 2010; Baldwin et al.,
320 2009). Five AgCht5 genes are clustered together in chromosome 2R with no other
321 intervening genes. The minimum distance is 340 bp between two chitinase ORFs and the
322 maximum is 2045 bp, suggesting that these genes may be derived from tandem duplications.
323 Gene duplication and loss according to a birth-and-death model of evolution is a feature of
324 the evolutionary history of the family 18 (GH18) of chitinases (Funkhouser et al., 2007).
325 Based on above information, we proposed that AgCht5 gene cluster may evolve primarily
326 from tandem duplication.
327 Because the transcripts of all the five Ag*Cht5* genes can be detected at various
328 developmental stages in *An*. *gambiae* (Fig. 5), these genes are transcribed and appear to be
329 independently regulated. However, the deduced AgCht5-2, AgCht5-3, AgCht5-4 and
330 AgCht5-5 proteins lack an essential catalytic glutamic acid (Watanabe et al., 1994, Lu et al.,
331 2002, Zhang et al., 2002, Zhu et al., 2008a). Therefore, they are presumed to act as
332 carbohydrate-binding proteins or lectins rather than as enzymes because they may not have
333 any catalytic activity (Goormachtig et al., 2001). Nevertheless, it should be pointed out that a

chitinase-like protein without catalytic activity may still play an important role in insect development. For example, insect imaginal disc growth factors (IDGFs) are chitinase-like proteins that are structurally related to chitinases but do not possess enzymatic activity. However, TcIDGF4 identified in *T. castaneum* might be involved in cell proliferation and contributed to adults ecdysis (Zhu et al., 2008b). Further studies will be necessary to elucidate the biological function of each of the five duplicated *Cht5* genes in *An. gambiae*.

Acknowledgements

Authors thank Sharon R. Starkey for maintaining mosquito colonies. This research was supported in part by Kansas Agricultural Experiment Station and K-State Arthropod Genomics Center funded by K-State Targeted Excellence program at Kansas State University. This manuscript is contribution No. 11-194-J from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas, USA. The *Anopheles gambiae* voucher specimens (voucher No. 211) are located in the Kansas State University Museum of Entomological and Prairie Arthropod Research, Manhattan, Kansas.

References

Ahmad, T., Rajagopal, R., Bhatnagar, R.K., 2003. Molecular characterization of chitinase from polyphagous pest *Helicoverpa armigera.* Biochem. Biophys. Res. Comm. 310, 188-195.

Ai, J., Yu, Q., Cheng, T., Dai, F., Zhang, X., Zhu, Y., Xiang Z., 2010. Characterization of multiple *CYP9A* genes in the silkworm *Bombyx mori*. Mol. Biol. Rep. 37, 1657-1664.

Arakane, Y., Muthukrishnan, S., 2010. Insect chitinase and chitinase-like proteins. Cell. Mol. Life Sci. 67, 201–216.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., 2003. Properties of catalytic, linker and chitin-binding domains of insect chitinase. Insect Biochem. Mol. Biol. 33, 631–648.

Baldwin, W.S., Marko, P.B., Nelson D.R., 2009. The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex.* BMC Genomics 10, 169.

Bolognesi, R., Arakane, Y., Muthukrishnan, S., Kramer, K.J., Terra, W.R., Ferreira, C., 2005. Sequences of cDNAs and expression of genes encoding chitin synthase and chitinase in the midgut of *Spodoptera frugiperda*. Insect Biochem. Mol. Biol. 35, 1249-1259.

Boot, R.G., Blommaart, E.F., Swart, E., Ghauharali-van der Vlugt, K., Bijl, N., Moe, C., Place, A., Aerts, J.M., 2001. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. J. Biol. Chem. 276, 6770-6778.

Filho, B. P.D., Lemos, F. J.A., Secundino, N. F.C., Pascoa V., Pereira, S. T., Pimenta, P. F.P., 2002. Presence of chitinase and beta-N-acetylglucosaminidase in the *Aedes aegypti*. a

chitinolytic system involving peritrophic matrix formation and degradation. Insect Biochem. Mol. Biol. 32, 1723-1729.

Fitches, E., Wilkinson, H., Bell, H., Bown, D.P., Gatehouse, J.A., Edwards, J.P., 2004. Cloning, expression and functional characterization of chitinase from larvae of tomato moth (*Lacanobia oleracea*): a demonstration of the insecticidal activity of insect chitinase. Insect Biochem. Mol. Biol. 34, 1037-1050.

Funkhouser, J.D., Aronson, N.N. Jr., 2007. Chitinase family GH18: Evolutionary insights from the genomic history of a diverse protein family. BMC Evol. Biol. 7, 96.

Genta, F. A., Blanes, L., Cristofoletti, P.T., do Lago, C. L., Terra W. R. and Ferreira, C., 2006. Purification, characterization and molecular cloning of the major chitinase from *Tenebrio molitor* larval midgut. Insect Biochem. Mol. Biol. 36, 789-800.

Goormachtig, S., Van de Velde, W., Lievens, S., Verplancke, C., Herman, S., De Keyser, A., Holsters, M., 2001. Srchi24, a chitinase homolog lacking an essential glutamic acid residue for hydrolytic activity, is induced during nodule development on *Sesbania rostrata.* Plant Physiol. 127, 78-89.

Gopalakrishnan, B., Muthukrishnan, S., Kramer, K. J., 1995. Baculovirus-mediated expression of a *Manduca sexta* chitinase gene: Properties of the recombinant protein. Insect Biochem. Mol. Biol. 25, 255-265.

Hirose, T., Sunazuka, T., Omura, S., 2010. Recent development of two chitinase inhibitors, Argin and Argadin, produced by soil microorganisms. Proc. Jpn. Acad., Series B 86, 85-102.

401 Khajuria, C., Buschman, L.L., Chen,M.-S., Muthukrishnan, S., Zhu, K.Y., 2010. A
402 gut-specific chitinase gene essential for regulation of chitin content of peritrophic
403 matrix and growth of *Ostrinia nubilalis* larvae. Insect Biochem. Mol. Biol. 40, 621-629.
404 Kim, M.G., Shin, S.W., Bae, K.S., Kim, S.C., Park, H.Y., 1998. Molecular cloning of
405 chitinase cDNAs from the silkworm, *Bombyx mori* and the fall webworm, *Hyphantria*
406 *cunea.* Insect Biochem. Mol. Bio. 28, 163-171.
407 Kong, H., Landherr, L.L., Frohlich, M.W., Leebens-Mack J., Ma, H. and dePamphilis, C.W.,
408 2007. Patterns of gene duplication in the plant *SKP1* gene family angiosperms:
409 Evidence for multiple mechanisms of rapid gene birth. Plant J. 50, 873-885.
410 Kramer, K.J., Corpuz, L., Choi, H.K., Muthukrishnan, S., 1993. Sequence of a cDNA and
411 expression of the gene encoding epidermal and gut chitinases of *Manduca sexta.* Insect
412 Biochem. Mol. Biol. 23, 691-701.
413 Krishnan, A., Nair, P.N., Jones, D., 1994. Isolation, cloning, and characterization of new
414 chitinase stored in active form in chitin lined venom reservoir. J. Biol. Chem. 269,
415 20971-20976.
416 Lu, Y., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and
417 functional analysis of active site acidic amino acid residues D142, D144 and E146 in
418 *Manduca sexta* (tobacco hornworm) chitinase. Insect Biochem. Mol. Biol. 32,
419 1369–1382.
420 Reese, T.A., Liang, H.E., Tager, A.M., Luster, A.D., Rooijen, V.N., Voehringer, D.,
421 Locksley, R.M., 2007. Chitin induces accumulation in tissue of innate immune cells
422 associated with allergy. Nature Let. 447, 92-96.

423 Reynolds, S. E. and Samuels, R. I., 1996. Physiology and biochemistry of insect moulting
424 fluid. Adv. Insect Physiol. 26, 158-232.
425 Royer, V., Fraichard, S., Bouhin, H., 2002. A novel putative insect chitinase with multiple
426 catalytic domains: Hormonal regulation during metamorphosis. Biochem. J. 366,
427 921-928.
428 Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture
429 research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U.S.A. 95,
430 5857-5864.
431 Shen, Z., Jacobs-Lorena, M., 1997. Characterization of a novel gut specific chitinase gene
432 from the human malaria vector *Anopheles gambiae*. J. Biol. Chem. 272, 28895-28900.
433 Shinoda, T., Kobayashi, J., Matsui, M., Chinzei, Y., 2001. Cloning and functional expression
434 of a chitinase cDNA from the common cutworm, *Spodoptera litura*, using a
435 recombinant baculovirus lacking the virus-encoded chitinase gene. Insect Biochem. Mol.
436 Biol. 1, 521-532.
437 Shuhui, L., Mok, Y.K., Wong ,W.S.F., 2009. Role of Mammalian Chitinases in Asthma. Int.
438 Arch. Allergy Immunol. 149, 369-377.
439 Spindler-Barth, M., Blattner, R., Vorgias, C.E., Spindler, K.D., 1998. Inhibition of two
440 family 18 chitinases by various allosamidin derivatives. Pestic. Sci. 52, 47-52.
441 Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary
442 Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

443 Watanabe, T., Uchida, M., Kobori, K., Tanaka, H., 1994. Site-directed mutagenesis of the
444 Asp-197 and Asp-202 residues in chitinase A1 of *Bacillus circulans* WL-12. Biosci.
445 Biotechnol. Biochem. 58, 2283–2285.
446 Zhang, H., Huang, X., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed
447 mutagenesis and functional analysis of an active site tryptophan of insect chitinase.
448 Insect Biochem. Mol. Biol. 32, 1477–1488.
449 Zhang, J.Z., Zhu, K.Y., 2006. Characterization of a chitin synthase cDNA and its increased
450 mRNA level associated with decreased chitin synthesis in *Anopheles quadrimaculatus*
451 exposed to diflubenzuron. Insect Biochem. Mol. Bio. 36, 712-725.
452 Zheng, Y., Zheng, S., Cheng, X., Ladd, T., Lingohr, E.J., Krell, P.J., Arif, B.M., Retnakaran,
453 A., Feng, Q., 2002. A molt-associated chitinase cDNA from the spruce budworm,
454 *Choristoneura fumiferana.* Insect Biochem. Mol. Biol. 32, 1813-1823.
455 Zheng, Y.P., Retnakaran, A., Krell, P.J., Arif, B.M., Primavera, M., Feng Q.L., 2003.
456 Temporal, spatial and induced expression of chitinase in the spruce budworm,
457 C*horistoneura fumiferana.* J. Insect Physiol. 49, 241-247.
458 Zhu, Q., Arakane Y, Banerjee, D., Beeman, RW, Kramer, K.J., Muthukrishnan, S., 2008a.
459 Domain organization and phylogenetic analysis of the chitinase like family of proteins
460 in three species of insects. Insect Biochem. Mol. Biol. 38, 452-466.
461 Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008b. Functional
462 specialization among insect chitinase family genes revealed by RNA interference.
463 Proc. Natl. Acad. Sci. U.S.A. 105, 6650-6655.

Zhu, Q., Deng, Y., Vanka, P., Brown, S.J., Muthukrishnan, S., Kramer, K.J., 2004. Computational identification of novel chitinase-like proteins in the *Drosophila melanogaster* genome. Bioinformatics 20, 161-169.

Figure legends:

Fig. 1. Genome structure, exon/intron organizations and protein domain architectures of five *AgCht5* genes. A) Genome structure of *AgCht5-1*, *AgCht5-2*, *AgCht5-3*, *AgCht5-4* and *AgCht5-5*: The light brown box and blue line represent DNA sequences of each gene and the linker region, respectively, between the ORF's. The length of linker region is marked under the blue line and by a yellow triangle. B) Exon and intron organizations: Exons are shown by green boxes whereas introns are shown by pink lines. C) Domain architectures of predicted AgCht5 proteins: Predicted signal peptide, catalytic domain and chitin-binding domain are boxed in yellow, gray and purple, respectively, whereas linker regions are shown by blue lines.

Fig. 2. Multiple alignments of deduced amino acid sequences of AgCht5's. Signal peptide, catalytic domain and chitin-binding domain are highlighted in orange, light blue and green, respectively, on the top of aligned sequences. The four conserved motif sequences are boxed in blue and denoted as CR1, CR2, CR3 and CR4. Fully conserved amino acid sequences are shaded in black.

Fig. 3. Comparative analysis of domain architecture of Cht5s from three mosquito species. The red line, blue triangle and light green hexagon represent the signal peptide, catalytic domain and chitin-binding domain, respectively.

Fig. 4. Phylogenetic analysis of catalytic domain sequences of putative chitinase 5 proteins from different species including *Aedes aegypti* (Aa, XP_001656234.1 for AaCht5-1, XP_001656233.1 for AaCht5-2, XP_001656232.1 for AaCht5-3, and XP_001656231.1 for AaCht5-4), *Anopheles gambiae* (Ag, HQ456129 for AgCht5-1, HQ456130 for AgCht5-2, HQ456131 for AgCht5-3, HQ456132 for AgCht5-4, and HQ456133 for AgCht5-5), *Bombyx mandarina* (Bma, AAG48700.1), *Bombyx mori* (Bmo, AAB47538), *Choristoneura fumiferana* (Cf, AAM43792), *Culex quinquefasciatus* (Cq, XP_001863384.1 for CqCht5-1, XP_001863385.1 for CqCht5-2, and XP_001863386.1 for CqCht5-3,) *Drosophila melanogaster* (Dm, CG9307), *Helicoverpa armigera* (Ha, AAQ91786), *Hyphantria cunea* (Hc, AAB47537), *Lacanobia oleracea* (Lo, CAF05663), *Manduca sexta* (Ms, P36362)，*Spodoptera frugiperda* (Sf, AAS18266), *Spodoptera litura* (Sl, AB032107), and *Tribolium castaneum* (Tc, AY675073). The phylogenetic tree was constructed using Mega 4 software (Tamura et al., 2007). Bootstrap values are obtained by the neighbor-joining method using 5000 replications. Bootstrap values are indicated only when greater than 40%.

Fig. 5. The expression patterns of five *AgCht5* genes in *Anopheles gambiae* as evaluated using RT-PCR. A) Gene expression patterns in eggs (EG), first- (L1), second- (L2), third- (L3) fourth- (L4) and fifth-instar larvae (L5); and adults (AD). B) Gene expression patterns in 12-, 24-, 36-, 48- and 60 h-old eggs as shown by EG12, EG24, EG36, EG48, and EG60, respectively. C) Gene expression patterns in 0-, 10-, 20-, 30- and 34-h-old eggs as shown by PU00, PU10, PU20, PU30 and PU34, respectively. *AgRps3* was used as reference gene for RT-PCR analysis.

Table 1. Sequences of PCR primers used in expression analyses of five *AgCht5* genes and 3'RACE primers used in amplifications of full-length cDNAs for *AgCht5-1* and *AgCht5-4* genes.

Primer name	Sequence (5'-3')	Product size (base pairs)
AgCht5-1-F	TTCCGGCTACAAGGACTTTG	188
AgCht5-1-R	TCGGGCTTTCGATCAGTTTC	
AgCht5-2-F	ACGATAAGGACAACTTTGTCTATC	152
AgCht5-2-R	GTCAGCACTCTCGCACAG	
AgCht5-3-F	GCTGTGTGAAATGCTGAAGG	166
AgCht5-3-R	TGCGTATATGCCACCCAATC	
AgCht5-4-F	TTCGCCAACCTGAAGAAGAC	146
AgCht5-4-R	TGGAGGAACTCAATCACACTG	
AgCht5-5-F	TTCATCGGCAGCGTGATC	197
AgCht5-5-R	TCGACCGGCACCTGTATC	
AgCht5-1-3'RACE-N1	ACGAGGACGAACGGTCGCTCCAGCAC	900
AgCht5-1-3'RACE-N2	TGGACGATTTCCACGGTCTTTGCGGGCCG	780
AgCht5-4-3'RACE-N1	ACTAGCAAACAGCGAGGAGCATGGACTG	680
AgCht5-4-3'RACE-N2	ACCGAAATGCAGCAGTCCGGCTGGGAG	520

553 Table 2. Number of deduced amino acid residues, presence or absence of a chitin-binding domain, availability of expressed sequence tag (EST)

554 in the NCBI database and genome location of each *AgCht5* gene from *An. gambiae*.

555

Gene	Amino acid residue	Catalytic domain	Chitin-binding domain	Availability of EST∗	Genome location
AgCht5-1	571	Yes	Yes	Yes	chr2R:21,584,333 - 21,587,318
AgCht5-2	412	Yes	No	Yes	chr2R:21,582,374 - 21,583,826
AgCht5-3	413	Yes	No	Yes	chr2R:21,578,829 - 21,580,211
AgCht5-4	409	Yes	No	Yes	chr2R:21,576,773 - 21,578,085
AgCht5-5	446	Yes	No	Yes	chr2R:21,573,544 - 21,574,884

556

557 ∗ Based on the *Anopheles gambiae* EST database from NCBI

Table 3. Percent identities of amino acid residues (nucleotides) among the ORF's of five *AgCht5* chitinase-like genes from *An. gambiae*.

	AgCht5-2	***AgCht5-3***	***AgCht5-4***	***AgCht5-5***
AgCht5-1	43 (58)	38 (52)	40 (55)	41 (57)
AgCht5-2		51 (59)	53 (63)	50 (66)
AgCht5-3			47 (57)	46 (58)
AgCht5-4				45 (58)

Fig. 1

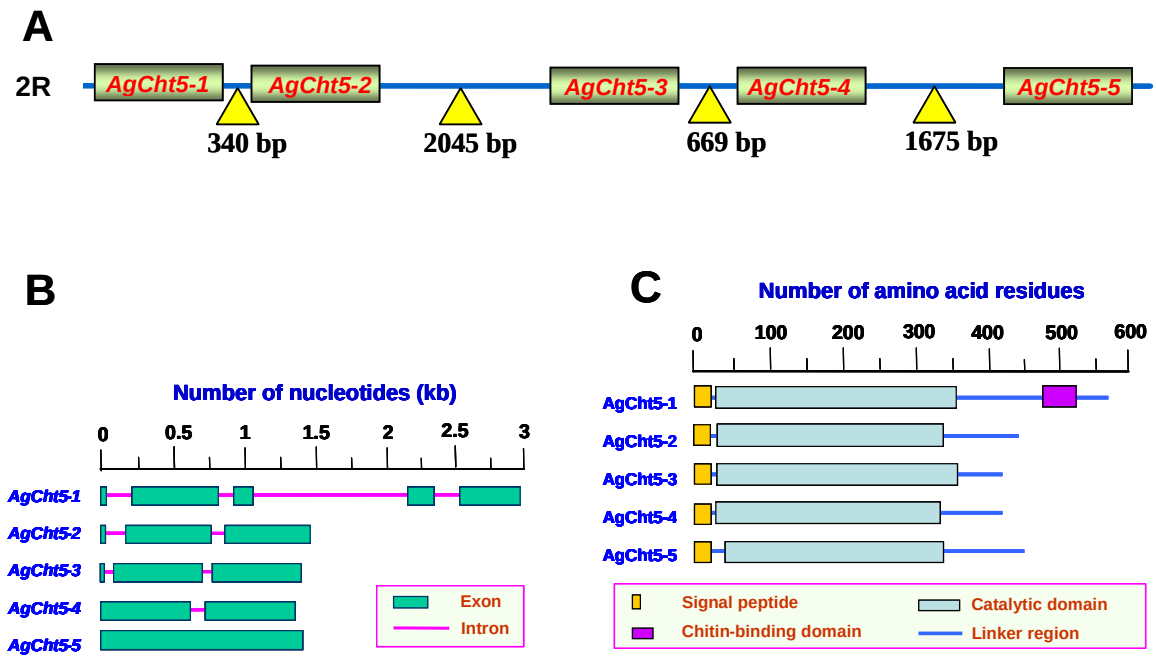

A
2R
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
340 bp
2045 bp
669 bp
1675 bp
B
Number of nucleotides (kb)
0
0.5
1
1.5
2
2.5
3
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
Exon
Intron
C
Number of amino acid residues
0
100
200
300
400
500
600
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
Signal peptide
Chitin-binding domain
Catalytic domain
Linker region

Fig. 2

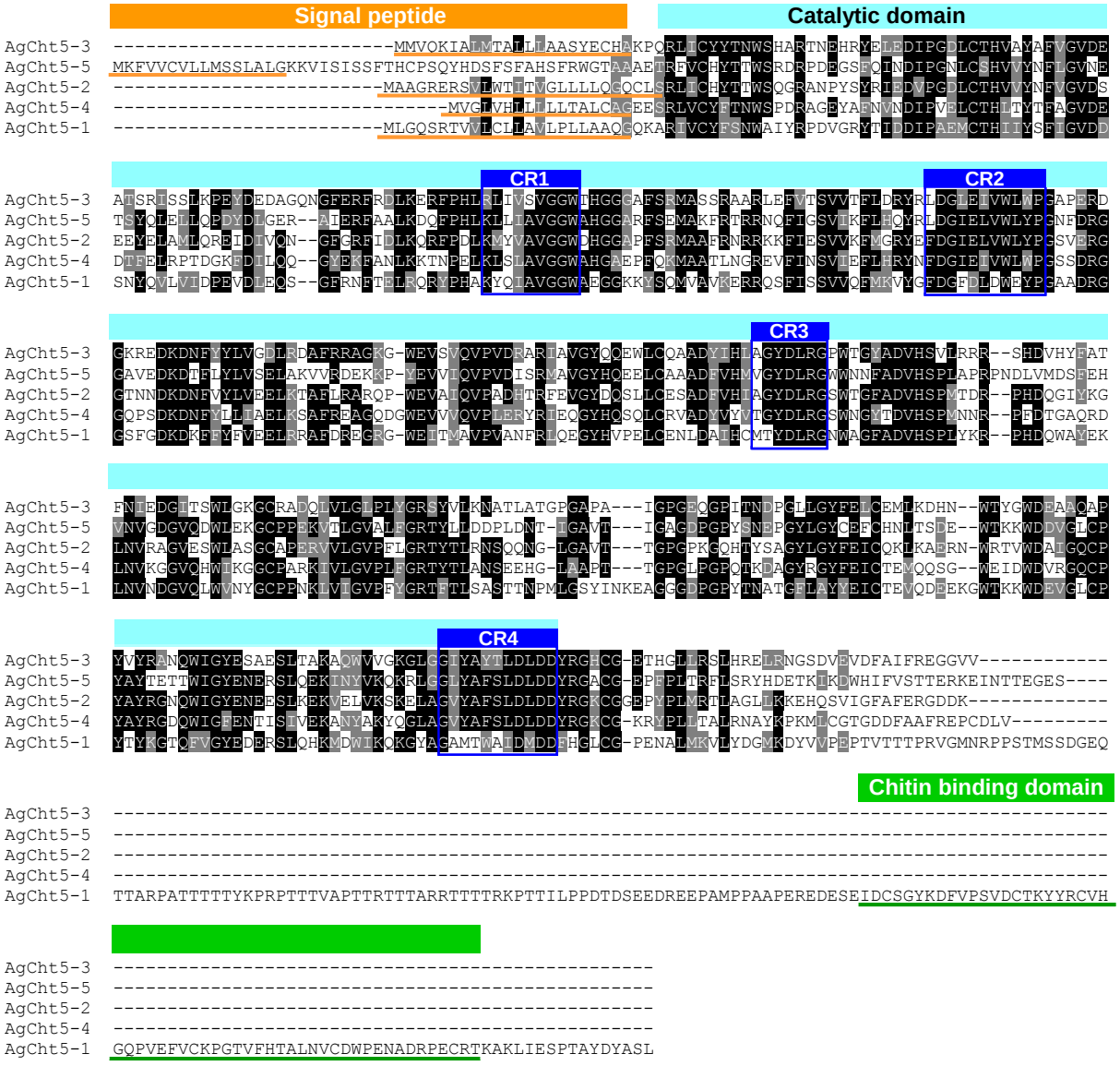

Signal peptide
Catalytic domain
CR1
CR2
CR3
CR4
Chitin binding domain
AgCht5-3
AgCht5-5
AgCht5-2
AgCht5-4
AgCht5-1

Fig. 3

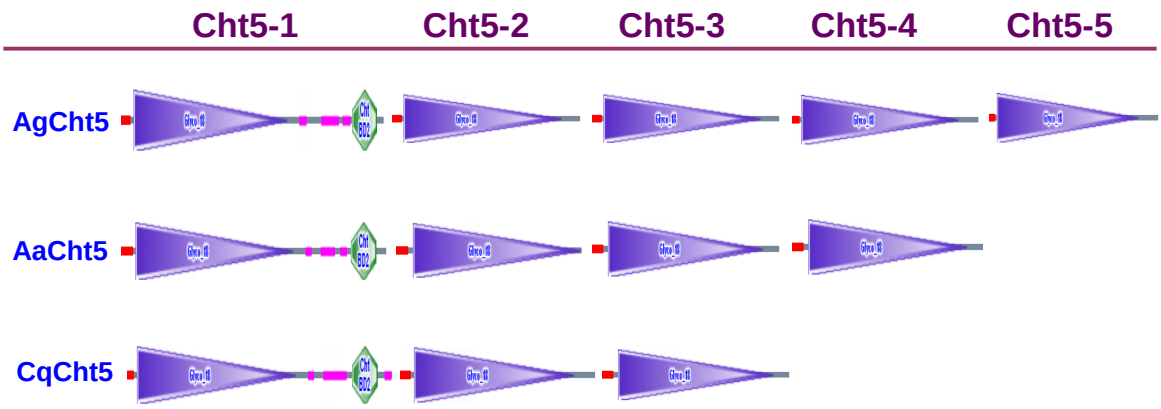


Fig. 4

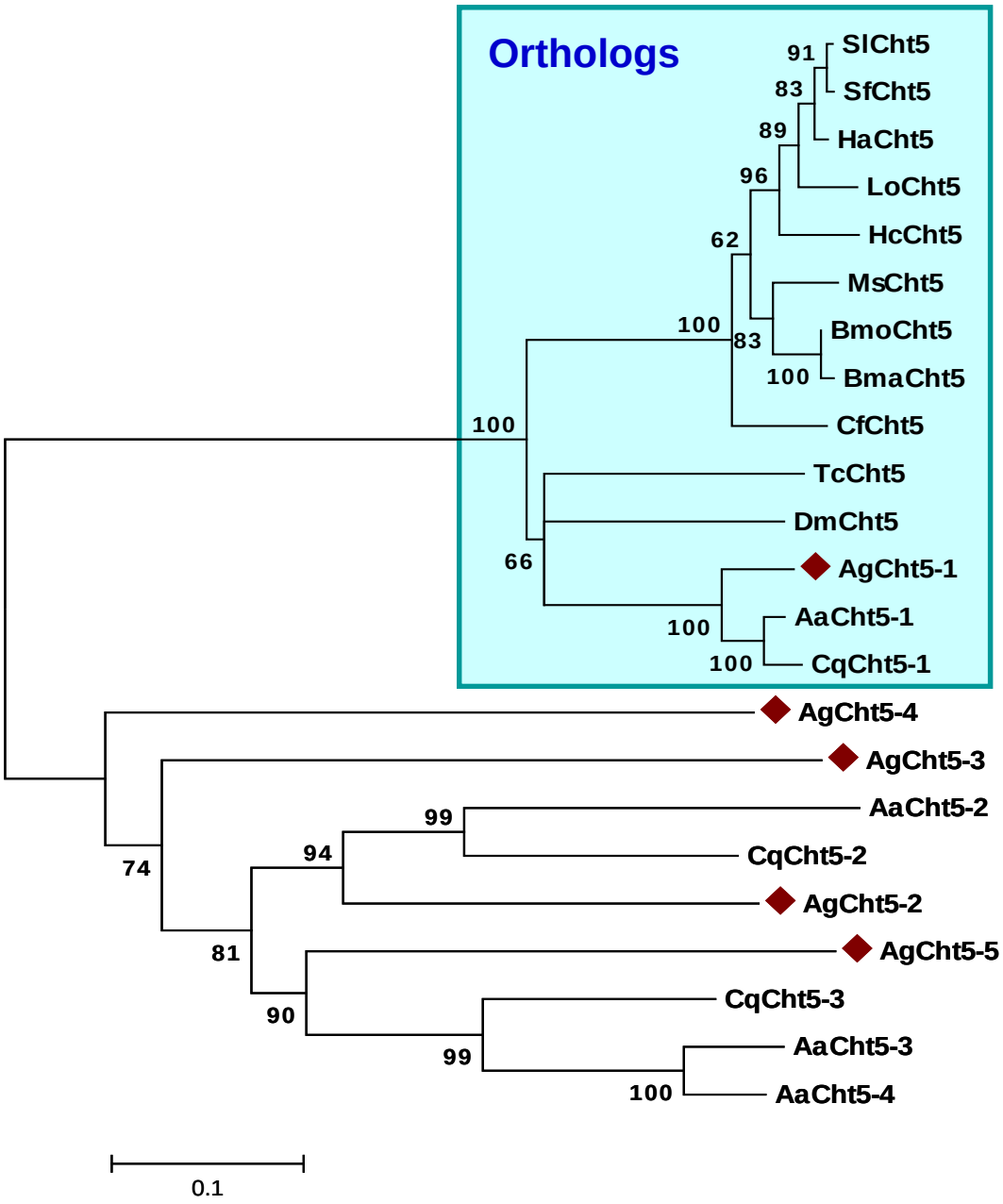
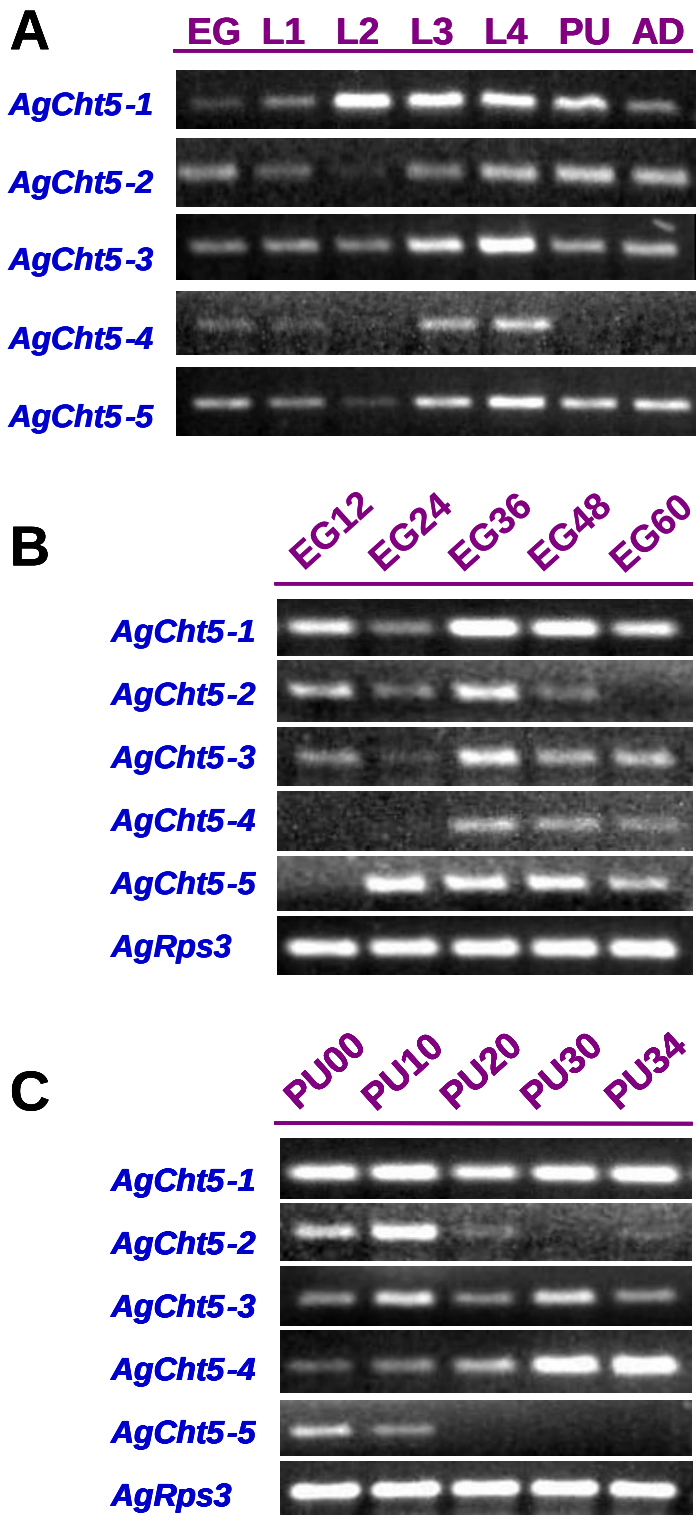

Orthologs
SlCht5
SfCht5
HaCht5
LoCht5
HcCht5
MsCht5
BmoCht5
BmaCht5
CfCht5
TcCht5
DmCht5
AgCht5-1
AaCht5-1
CqCht5-1
AgCht5-4
AgCht5-3
AaCht5-2
CqCht5-2
AgCht5-2
AgCht5-5
CqCht5-3
AaCht5-3
AaCht5-4
91
83
89
96
62
100
83
100
100
66
100
100
99
94
74
81
90
99
100
0.1

Fig. 5


A
EG L1 L2 L3 L4 PU AD
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
B
EG12 EG24 EG36 EG48 EG60
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
AgRps3
C
PU00 PU10 PU20 PU30 PU34
AgCht5-1
AgCht5-2
AgCht5-3
AgCht5-4
AgCht5-5
AgRps3